Bridging Task Expressions and Search Queries

Chirag Shah
University of Washington
Seattle, WA, USA
chirags@uw.edu

Ryen W. White
Microsoft Research
Redmond, WA, USA
ryenw@microsoft.com

ABSTRACT

People engage in search episodes as they have a task or a problematic situation. Often this task is not clearly expressed by the information seeker, nor directly supported by the search system. People also list their tasks using tools such as to-do applications, and while many of these could be search tasks, there is a lack of that recognition or a possible bridge to a search system. In the work reported here, we aim to create that bridge by analyzing data on both sides. In task management, we examined 1,000 to-do tasks annotated by human assessors for their appropriateness for a search engine and created a simple process to learn that classification. In search, we analyzed millions of queries in a search engine log to understand how often queries represent tasks that people express in to-do lists. Our results show that (1) we can accurately predict which of the to-do tasks are appropriate as search queries; and (2) such tasks do indeed show up in search engines as a substantial segment. Together, these findings outline an opportunity to link explicitly expressed tasks to search queries and vice versa. This has implications for both task completion and query understanding.

CCS CONCEPTS

• **Information systems → Retrieval tasks and goals**; **Task models**.

KEYWORDS

Search tasks; queries; log analysis; human assessment

ACM Reference Format:
Chirag Shah and Ryen W. White. 2021. Bridging Task Expressions and Search Queries. In *Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21), March 14–19, 2021, Canberra, ACT, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3406522.3446045

1 INTRODUCTION

Despite the importance of user tasks, most search systems are still disconnected from them. Some scholars have attempted to address this by deriving task information from user behaviors [6, 8, 15, 17, 18]. While recognizing implicit tasks is an important avenue for research [21, 22], we overlook the fact that people do express their tasks explicitly – only that this happens outside the search system – typically in to-do applications such as Google Keep, Todoist, and Microsoft To Do. We ask the question: can we analyze those tasks and identify those that are appropriate for search systems? And if we can, what is the likelihood that the task expression (typically a short phrase such as 'order contact lens') can map directly to a search query? We set out to answer these questions to bridge the explicit expression of to-do tasks and web search.

For this, we used aggregated tasks data from a popular to-do application, deployed to millions of users. From this, we selected 1,000 tasks and used human assessment to identify those appropriate for web search. We then devised a simple algorithm to perform such classification automatically. Finally, we analyzed search log data to verify that such tasks indeed do appear in search engines. Our findings are the first steps in bridging task listings and searching.

2 BACKGROUND

In the domain of information retrieval (IR), a task is considered as a set of connected physical, affective, and cognitive actions through which individuals try to accomplish goals in their work or everyday lives [3, 26]. It is an expression or representation of the *goal* or *purpose* of the search process (e.g., "gathering information to write a research report" or "planning a trip") [11]. From a more granular perspective, a task can be interpreted as an atomic information need resulting in one or more search queries [10]. According to existing empirical work, tasks are inherently hierarchical, multidimensional, and modeled and interpreted at multiple granularity levels [4, 5].

In the context of search-related tasks, scholars have considered three levels: work task, information seeking task, and search or retrieval task, each with its goals, intentions, conditions, actions, and outcomes [4]. At its most general level, work task refers to the broad, overarching goal which triggers the search process. That task comprises one or more information seeking tasks, which could be accomplished through multiple consultations with information sources such as search systems and human experts. Each seeking task can then be further deconstructed to one or more search tasks that can be accomplished through a single consultation with search systems [5, 19, 30]. Each level of task has a defined goal but an anticipated and likely unknown outcome [25]. Furthermore, each sub-task level is recursively dependent on the previous one(s) [19].

Several studies have found varied task types based on various behavioral signals such as the number of content pages and queries [9, 16, 20], the time spent on pages and queries [9, 12, 16, 20], click and scroll depth [9], query reformulations [6, 9, 16], and task completion time [9]. These studies have used these data to determine the relationship between task and search behaviors, and try to explicate task characteristics (e.g., goal or product of the task).

Studies of to-do tasks, on the other hand, have focused on task management, including how people plan and organize their tasks

CHIIR '21, March 14–19, 2021, Canberra, ACT, Australia

ACM ISBN 978-1-4503-8055-3/21/03…$15.00
https://doi.org/10.1145/3406522.3446045

[2]. The recent availability of large-scale logs has enabled methods such as task duration estimation [27], task completion detection [28], and enhancing notifications to maximize completion [7]. Systems have supported task completion by allowing users to focus on tasks requiring human intervention [1, 23] or by generating action plans [14]. Stumpf et al. [24] and Kiseleva et al. [13] explore methods to understand people's task intent and provide the resources required to complete those tasks, albeit not focused on to-do tasks and the connection between such tasks and search.

In this paper, we focus on the tasks people express in their to-do applications and seek to connect those to search tasks, with a goal that they could be executed directly on a search engine. This area, at the intersection of task management and search, is unexplored and untapped, with a lot of potential for users and system designers.

3 DATA

Tasks in to-do lists can be divided into two broad categories: actionable and non-actionable. An actionable task is something that one can readily act on. Examples include 'buy a backpack', 'find a new doctor', and 'milk' in a grocery list. Non-actionable tasks are typically lists of things that serve as reminders or a way of notetaking. Examples include a list of books to read, and things to bring on a camping trip. They also include intangible, vague, or overly broad tasks, such as 'make a difference'. For the purpose of the experiments reported here, we focus on actionable tasks only.

We started with an anonymized, aggregated subset of tasks appearing in the, now-defunct, Wunderlist application. Tasks were stripped of any personally identifiable information (proper names, phone numbers, etc.) before analysis. The application had a default list, to which tasks were added if the user did not specify a list. We wanted to focus on tasks where the intent was clear from the task titles alone. Tasks titles in non-default lists often rely on the list name as context, e.g., the 'milk' example in the grocery list, where the intent is clearly to buy milk even if not stated explicitly. From the default list, we needed to filter out the tasks that either did not have a leading verb that indicate an action and/or the context was very clear. For example, 'call mom' would not be considered because it was clearly a task that had an assigned context (phone). We sorted these tasks in descending order by their frequency to identify top action verbs in the default list. We reviewed the top 50 most frequent tasks and selected verbs that could have a context outside of certain strongly indicative action terms such as phone/call, email, print, and clean. The final set of verbs considered for identifying suitable tasks for this dataset were: *book, bring, buy, change, check, collect, find, finish, fix, get, look, make, move, order, pay, pick, post, put, renew, reschedule, return, sell, send, set, submit, take, update.*

Using this subset of selected action verbs, we filtered the aggregated tasks data and identified a set of tasks that (1) started with one of these verbs; (2) were added to the default list; and (3) were found in the lists for at least 100 users. This generated a total of 7,563 tasks. We sorted these in descending order by user frequency and selected the top 1,000 for human annotation.

4 ANNOTATIONS

We used a proprietary crowdsourcing service to get task annotations from human judges. Judges were given the following instructions.

> *People use To-do apps (e.g., Todoist, Google Keep, Microsoft To Do) to list things they want to get done. Some of these are to be done in physical spaces (home, office, store), whereas others can be done online, at least partially. And some of the online tasks could benefit from the use of a search engine. We are interested in learning how you would do the following tasks.*
>
> *Definition of an online task without search engine: A to-do item that will need at least some online service (a website), but not require searching. For example, one could pay a utility bill by just going to a website and not doing any searching.*
>
> *Definition of an online task with search engine support: A to-do item that could use at least some amount of searching. For example, 'look for a chiropractor', 'find a babysitter', and 'buy a backpack' can be done using a search engine.*
>
> *In contrast, 'look for keys', 'email Bob', and 'call Susan' can be done without searching or even going online.*
>
> *With this in mind, please pick an appropriate option based on how you would do the given task. If you pick I'm not sure / I don't understand this task, please leave a comment explaining why.*

Multiple rounds of internal testing were performed before sharing the full set of 1,000 tasks with the judges. The judges were first given a separate set of 100 tasks for pilot testing. Their feedback was incorporated in the instructions and data presentation. To improve the robustness of the labeling, we used multiple judges per task. Five different judges provided judgments (ratings and optional comments) for the full set of 1,000 tasks.

5 ANALYSIS

As we analyzed the judgments for the 1,000 tasks, we found that some tasks clearly fit one of the four categories (offline, online, search, and unclear), whereas some of them had a majority label (3 or 4 out of 5 judges picked the same category). Out of 5,000 total judgments, only 20 were marked as 'Unclear'. In almost all of the cases, the reason was insufficient information or context. The task 'book review' was one where all five judges marked it as 'Unclear'. Two other tasks worth noticing were 'book group' (4/5 marking as 'Unclear') and 'check out' (3/5 marking as 'Unclear').

We then investigated the aggregated judgments for the tasks, grouped by the leading verbs, e.g., we found that when a task starts with 'buy', 86% of the time judges marked it for 'search'. We also studied how often the tasks with the leading verb gets the label using majority voting, e.g., tasks starting with 'book' get the 'Online' label 76% of the time through majority voting. Table 1 presents the results for all verbs. Here, 'Aggregated label' and $P(A)$ indicate overall what was the most common label for all tasks starting with the given verb and the corresponding probability. Similarly, 'Majority label' and $P(M)$ indicate how often tasks starting by the given verb received a label as picked by a majority of the judges.

Table 1: Task starting verbs with their labels by either aggregating or majority vote, and corresponding probabilities.

Verb	Agg. label	$P(A)$	Maj. label	$P(M)$
book	online	0.72	online	0.76
bring	offline	0.98	offline	1.00
buy	search	0.86	search	0.86
change	offline	0.50	offline	0.51
check	online	0.55	online	0.62
collect	offline	1.00	offline	1.00
find	search	0.74	search	0.71
finish	offline	0.54	offline	0.50
fix	offline	0.90	offline	0.88
get	offline	0.60	offline	0.54
look	search	0.75	search	0.75
make	offline	0.62	offline	0.63
move	offline	0.90	offline	0.86
order	search	0.52	search	0.66
pay	online	0.98	online	0.98
pick	offline	0.96	offline	1.00
post	offline	0.55	offline	0.69
put	offline	1.00	offline	1.00
renew	online	0.97	online	1.00
reschedule	online	0.96	online	1.00
return	online	0.76	online	0.85
sell	online	0.47	online	0.64
send	online	0.82	online	0.85
set	offline	0.67	offline	0.63
submit	online	0.97	online	1.00
take	offline	1.00	offline	1.00
update	online	0.89	online	0.95

We found that for 16 out of 27 verbs considered here, we could potentially predict the label for a task starting with that verb with high confidence. Those that were most unclear ($P < 0.5$) were 'change' and 'sell'. The confusion for 'change' tasks was between offline and online and depended on the noun following 'change', e.g., 'change bed' meant offline, whereas 'change dentist appointment' meant online. Based on this, we hypothesized that if the noun is tangible (bed, oil filter), the task is 'offline', whereas if the noun is abstract (password, appointment, address), the task is 'online'.

The confusion for 'sell' tasks was online vs. search, related to the unclear utility of search in selling something. Interestingly, while 'buy' tasks were clearly marked for search, only 52% of times 'order' tasks got a majority for search. Here also, the confusion lay in online vs. search. This may be due to the perceived intentions behind these two verbs: 'buy' may be perceived as tentative and thus benefiting from search, whereas 'order' may indicate a resolved intention that may or may not require a search engine. This hypothesis can easily be tested by having judges annotate the same noun with a different starting verb, e.g., 'buy dog food' and 'order dog food'.

6 A SIMPLE RULE-BASED ALGORITHM FOR TASK LABELING

Based on our analysis and observations, we devised the following simple algorithm to assign a task 'offline', 'online', or 'search' labels.

Algorithm 1 : Task Classification.

Input: Task titles/phrases
Output: Category label

Initialize label='unknown'
if the task starts with verbs *collect, put, take, bring, pick, fix, move, set, make, get, finish* **then**
  label='offline'
end if
if the task starts with verbs *pay, renew, submit, reschedule, update, send, return, book, check, sell* **then**
  label='online'
end if
if the task starts with verbs *buy, look, find, order* **then**
  label='search'
end if
if the task starts with verbs *change*, followed by a tangible noun **then**
  label='online'
end if
if the task starts with verbs *post*, followed by an online service **then**
  label='online'
else
  label='offline'
end if

Applying this algorithm to our dataset of 1,000 tasks, we obtain 82% accuracy in task classification if we consider aggregated labels as the ground truth. If we consider majority labels as ground truth, the accuracy increases to 84%. Note that if we do further refinement of our crude rules (e.g., more nuanced handling of the 'order' tasks, as described above), we could potentially obtain higher accuracies. Additional analysis of the false positive and false negative cases could give us a better understanding of which of the tasks are being misclassified and perhaps why this is happening.

7 CONNECTING TASKS WITH SEARCHES

Next, we examined how the actionable tasks from Wunderlist show up in searches to the Bing search engine. Specifically, we took those 1,000 tasks and looked for exact matches in Bing search logs from May 1-7, 2020, June 1-7, 2020, and July 1-7, 2020. The 'Original' columns in Table 2 present the results. Here, the 'Tasks' column indicates how many tasks were in each of the three categories. 'Query freq' is the aggregated frequency of queries that match with the tasks. 'Avg freq' is 'Query freq' divided by 'Tasks'.

As we can see, the offline tasks do not appear as often in the search logs. But search-friendly tasks also do not seem to appear often either. This may be because people do not enter their to-do tasks verbatim in search engines. For instance, for a to-do task 'buy bus tickets', the corresponding query would likely be 'bus tickets'.

When we match the tasks with the search queries after removing 'buy', 'order', and 'find' from the task titles, we obtain a huge spike in search-friendly task matches. The new statistics are shown in the 'Preceding verbs removed' columns in Table 2.

Table 2: Frequencies of tasks in different categories as represented in the Bing query log over a three-week (non-contiguous) time period. Tasks are matched to queries in their original forms, with preceding verbs removed, and 'amazon' tasks removed.

		Original		Preceding verbs removed		'Amazon' tasks removed	
Category	**Tasks**	**Query freq**	**Avg freq**	**Query freq**	**Avg freq**	**Query freq**	**Avg freq**
Offline	405	8,932	22.11	8,932	22.11	8,932	22.11
Online	366	21,283	58.15	21,283	58.15	21,283	58.15
Search	229	6,390	27.90	9,815,280	42,861.48	2,237,024	9,768.66

When we review the actual matches, the top one is 'amazon' appearing millions of times in the logs. The original task was 'order amazon', which appeared only a small number of times in the tasks data. However, since we removed 'order' for query matching, we were matching only 'amazon', a popular, navigational query. If we exclude the 'amazon' tasks as outliers in our analysis, we still obtain high query frequencies for the 'search' category (as shown in the 'Amazon tasks removed' columns of Table 2).

These results clearly show that tasks that our algorithm determined to be search-friendly are indeed very likely to show up in search engines. To further verify this, we ranked the matched queries by their frequency in descending order. Then, we computed reciprocal rank (RR) and weighted RR (WRR) for each of the three categories (online, offline, and search) as follows.

$$RR_{category} = \sum \frac{1}{rank_i} \quad \text{where label(i)} = category \tag{1}$$

$$WRR_{category} = \sum freq(i) \cdot \frac{1}{rank_i} \quad \text{where label(i)} = category \tag{2}$$

We then took the averages across all the queries in a given category for RR and WRR to compute mean values of MRR and MWRR. The results are in Table 3. Here, the higher MRR indicates the results belonging to that category are at higher ranks based on the frequency with which they appear in the search logs. It is clear that 'search' labeled tasks have queries matching much higher ranks than those with 'online' and certainly 'offline'. And when we weight these ranks with their corresponding frequencies using MWRR, it is abundantly clear that what we identified as 'search-friendly' tasks are indeed the highest ranked and most prominent queries among all task-related queries. This means the users are already using search engines for the tasks we identified as 'search-friendly' more so than for other types of tasks, providing another form of validation for our results.

Table 3: Mean reciprocal rank (MRR) and mean weighted MRR (MWRR) for queries with different category labels.

	MRR	**MWRR**
Offline	0.0019	0.1119
Online	0.0022	0.8401
Search	0.0256	2,826.32

8 ANALYZING UNMATCHED SEARCH TASKS

Not all tasks that were labeled suitable for search appeared in our Bing log sample. We examined such tasks to better understand why.

Of course, the simplest explanation is that we only took a sample of Bing logs (first seven days of each month from May to July 2020). Had we considered a longer duration or a larger sample, we may have found matches for all these tasks. Nonetheless, looking at the search tasks that did not find a match with any query, we found that out of 48 in this list, 32 were 'buy' tasks. It makes sense that they did not match any query because (as noted earlier) anyone looking to buy something will presumably not enter 'buy' with that item in a search engine. Nine items were 'order' tasks, five were 'find' tasks, and two were 'look at/for' tasks.

Once we removed 'buy', 'order', and 'find' from task titles and looked for query matches, we were left with only 25 unmatched search tasks. Almost all the missing items were actual objects that people did not search for in our sampled log. This can likely be attributed to the fact that people are not doing all their shopping tasks on a search engine as they visit specialized e-commerce services for them. A couple of exceptions were those 'look at/for' tasks. We believe 'look at/for' should also be removed when matching with search queries, as these tasks are clear candidates for searching.

9 CONCLUSION

We believe we have provided a thorough analysis and compelling evidence that we can identify search-friendly tasks from one's to-do list with great accuracy and reliability. While almost any task can be linked to a search engine, in order to gain user trust and help ensure the usability of this proposed functionality, it is important to provide such a link only for the tasks for which we have sufficient confidence. Considering this, we provide the following recommendation for the overall flow of such a "task-to-search" functionality:

(1) For a given item in one's to-do list, identify if it is an actionable item (as in [29]).
(2) For a given actionable item, use the algorithm given here to identify if it is suitable for offline, online, or search.
(3) If an item is deemed to be search-friendly, create a link to a search engine and provide it with the item on the interface.

In addition to continuing to work with these steps, we are also considering tasks requiring multiple queries by analyzing *sessions* instead of individual queries in the search logs. This becomes quite important as we acknowledge and address the fact that many tasks are not accomplished using single queries or even a collection of queries directly executed on a search engine. To truly address the underlying tasks, one needs to contextualize and perhaps expand its meaning and then consider how or if search can help. A part of this contextualization can also be achieved by considering various attributes around a task's expression such as time, place, other tasks entered at the same time, and a user's own background and profile.

REFERENCES

[1] Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: The short of mobile device use sessions. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 243–252.

[2] Victoria Bellotti, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G Bobrow, and Nicolas Ducheneaut. 2004. What a to-do: Studies of task management towards the design of a personal task list manager. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, 735–742.

[3] Katriina Byström. 2007. Approaches to "task" in contemporary information studies. In *Proceedings of the 6th International Conference on Conceptions of Library and Information Science*. 1–10.

[4] Katriina Byström and Preben Hansen. 2002. Work tasks as units for analysis in information seeking and retrieval studies. *Emerging frameworks and methods* (2002), 239–251.

[5] Katriina Byström and Preben Hansen. 2005. Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology* 56, 10 (2005), 1050–1061.

[6] Michael J Cole, Chathra Hendahewa, Nicholas J Belkin, and Chirag Shah. 2015. User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)* 33, 1 (2015), 1.

[7] David Graus, Paul N Bennett, Ryen W White, and Eric Horvitz. 2016. Analyzing and predicting task reminders. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. 7–15.

[8] Ahmed Hassan Awadallah, Ryen W White, Patrick Pantel, Susan T Dumais, and Yi-Min Wang. 2014. Supporting complex search tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 829–838.

[9] Jiepu Jiang, Daqing He, and James Allan. 2014. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 607–616.

[10] Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. 2006. Generating query substitutions. In *Proceedings of the 15th International Conference on the World Wide Web*. 387–396.

[11] Diane Kelly. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval* 3, 12 (2009).

[12] Diane Kelly and Nicholas J Belkin. 2004. Display time as implicit feedback: Understanding task effects. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 377–384.

[13] Julia Kiseleva, Hoang Thanh Lam, Mykola Pechenizkiy, and Toon Calders. 2013. Discovering temporal hidden contexts in web sessions for user trail prediction. In *Proceedings of the 22nd International Conference on the World Wide Web*. 1067–1074.

[14] Nicolas Kokkalis, Thomas Köhn, Johannes Huebner, Moontae Lee, Florian Schulze, and Scott R Klemmer. 2013. Taskgenies: Automatically providing action plans helps people complete tasks. *ACM Transactions on Computer-Human Interaction* 20, 5 (2013), 27.

[15] Zhen Liao, Yang Song, Li-wei He, and Yalou Huang. 2012. Evaluating the effectiveness of search task trails. In *Proceedings of the 21st International Conference on the World Wide Web*. 489–498.

[16] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. 2010. Search behaviors in different task types. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*. ACM, 69–78.

[17] Jiqun Liu, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. 2019. Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search. In *Proceedings of the 4th ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, 123–132.

[18] Jiqun Liu, Shawon Sarkar, and Chirag Shah. 2020. Identifying and predicting the states of complex search tasks. In *Proceedings of the 5th ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, 193–202.

[19] Gary Marchionini and Hermann Maurer. 1995. The roles of digital libraries in teaching and learning. *Commun. ACM* 38, 4 (1995), 67–75.

[20] Matthew Mitsui, Jiqun Liu, and Chirag Shah. 2018. The paradox of personalization: Does task prediction require individualized models?. In *Proceedings of the 3rd ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, 277–280.

[21] Matthew Mitsui and Chirag Shah. 2018. The broad view of task type using path analysis. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval*. ACM, 131–138.

[22] Matthew Mitsui and Chirag Shah. 2019. Bridging gaps: Predicting user and task characteristics from partial user information. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. ACM, 415–424.

[23] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah L McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe. 2007. An intelligent personal assistant for task and time management. *AI Magazine* 28, 2 (2007), 47–47.

[24] Simone Stumpf, Xinlong Bao, Anton Dragunov, Thomas G Dietterich, Jon Herlocker, Kevin Johnsrude, Lida Li, and J Shen. 2005. Predicting user tasks: I know what you're doing. In *Proceedings of the AAAI Workshop on Human Comprehensible Machine Learning*.

[25] Elaine G. Toms. 2011. Task-based information searching and retrieval. *Interactive Information Seeking, Behaviour and Retrieval* (2011), 43–60.

[26] Pertti Vakkari. 2003. Task-based information searching. *Annual Review of Information Science and Technology* 37, 1 (2003), 413–464.

[27] Ryen W White and Ahmed Hassan Awadallah. 2019. Task duration estimation. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. ACM, 636–644.

[28] Ryen W White, Ahmed Hassan Awadallah, and Robert Sim. 2019. Task completion detection: A study in the context of intelligent systems. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 405–414.

[29] Ryen W. White, Elnaz Nouri, James Woffinden-Luey, Mark Encarnacion, and Sujay Kumar Jauhar. 2020. Microtask detection. *ACM Transactions on Information Systems (TOIS)* in press (2020).

[30] Hong Xie. 1997. Planned and situated aspects in interactive IR: Patterns of user interactive intentions and information seeking strategies. In *Proceedings of the American Society for Information Science and Technology Annual Meeting*. 101–110.